GPU-Accelerated Genomic Sequence Alignment for Bioinformatics

Abill Robert

July 2, 2024

GPU-Accelerated Genomic Sequence Alignment for Bioinformatics

AUTHOR

ABILL ROBERT

DATA: June 25, 2024

Abstract

The exponential growth of genomic data has necessitated the development of efficient and scalable computational techniques for sequence alignment, a fundamental task in bioinformatics. Traditional CPU-based methods often struggle with the increasing volume and complexity of genomic sequences, leading to significant delays in data processing and analysis. This paper explores the utilization of Graphics Processing Units (GPUs) to accelerate genomic sequence alignment, leveraging their parallel processing capabilities to enhance performance and reduce computational time. We review current GPU-accelerated algorithms and frameworks, such as CUDA and OpenCL, highlighting their architecture and implementation strategies. Through a series of benchmark tests, we demonstrate the substantial improvements in alignment speed and efficiency achieved by GPU-based approaches compared to conventional CPU-based methods. Additionally, we discuss the challenges and limitations associated with GPU acceleration, including memory management and algorithm optimization, and propose potential solutions to address these issues. Our findings underscore the potential of GPU-accelerated genomic sequence alignment to transform bioinformatics workflows, enabling faster and more accurate analysis of large-scale genomic data, thereby facilitating advancements in personalized medicine, evolutionary biology, and other related fields.

Introduction

The field of bioinformatics has been revolutionized by the rapid advancements in sequencing technologies, leading to an unprecedented influx of genomic data. The ability to align genomic sequences accurately and efficiently is critical for a wide range of applications, including disease diagnosis, evolutionary studies, and personalized medicine. Traditional methods for genomic sequence alignment, which predominantly rely on Central Processing Units (CPUs), often face significant challenges in handling the sheer volume and complexity of modern genomic datasets. These methods can be time-consuming and computationally intensive, making them less feasible for large-scale genomic studies.

Graphics Processing Units (GPUs), originally designed for rendering graphics, have emerged as powerful tools for general-purpose computing due to their highly parallel architecture. Unlike CPUs, which are optimized for sequential processing, GPUs are capable of performing many

operations simultaneously, making them particularly well-suited for tasks that can be parallelized. This inherent parallelism offers a promising avenue for accelerating computationally intensive tasks in bioinformatics, such as genomic sequence alignment.

In recent years, there has been a growing interest in leveraging GPU acceleration to enhance the performance of sequence alignment algorithms. GPU-accelerated sequence alignment takes advantage of the parallel processing capabilities of GPUs to significantly reduce the time required for aligning large genomic datasets. Various algorithms and frameworks, such as CUDA (Compute Unified Device Architecture) and OpenCL (Open Computing Language), have been developed to harness the power of GPUs for bioinformatics applications.

This paper aims to explore the potential of GPU-accelerated genomic sequence alignment by providing a comprehensive overview of current techniques and their implementation. We will review the architecture and principles behind GPU computing, discuss existing GPU-accelerated alignment algorithms, and evaluate their performance through benchmark studies. Additionally, we will address the challenges and limitations associated with GPU acceleration, including issues related to memory management and algorithm optimization.

By highlighting the advantages and potential of GPU-accelerated genomic sequence alignment, this paper seeks to contribute to the ongoing efforts to improve computational efficiency in bioinformatics. The insights gained from this study may facilitate the development of faster and more accurate alignment tools, ultimately advancing research in genomics and related fields.

II. Background

1. Traditional CPU-Based Sequence Alignment Algorithms

Genomic sequence alignment is a critical task in bioinformatics, essential for identifying regions of similarity between sequences, which can provide insights into functional, structural, and evolutionary relationships. Several traditional algorithms have been developed to perform sequence alignment efficiently using Central Processing Units (CPUs). Notable among these are BLAST (Basic Local Alignment Search Tool) and Bowtie.

BLAST: BLAST is one of the most widely used tools for comparing an input sequence (query) against a database of sequences to find regions of local similarity. It employs a heuristic approach to speed up the alignment process by first finding short, high-scoring segment pairs (HSPs) between the query and database sequences. These HSPs are then extended to form alignments. BLAST's speed and relatively high accuracy have made it a staple in bioinformatics, but its performance degrades with the increasing size of genomic databases.

Bowtie: Bowtie is another popular tool designed for aligning short DNA sequences (reads) against large reference genomes. It uses a Burrows-Wheeler Transform (BWT) and a full-text minute-space (FM) index to achieve fast and memory-efficient alignment. Bowtie excels in processing high-throughput sequencing data, offering significant improvements over earlier methods. However, as sequencing technologies advance, producing vast amounts of data, even Bowtie's efficiency can become a bottleneck.

2. Limitations of CPU-Based Approaches

Despite the advancements in CPU-based alignment algorithms, they face significant limitations when dealing with the massive scale of modern genomic data. The primary challenges include:

Scalability: CPU-based algorithms struggle to scale efficiently with the increasing size and complexity of genomic datasets. The sequential nature of CPU processing limits the speed at which large-scale alignments can be performed.

Computational Intensity: Sequence alignment involves extensive computation, including searching, matching, and extending alignments. This intensity results in substantial processing time and power consumption, which can be prohibitive for large datasets.

Memory Bottlenecks: Large genomic databases require significant memory for storage and processing. CPUs, with their limited memory bandwidth and cache sizes, often encounter bottlenecks that slow down the alignment process.

3. Introduction to GPU Architecture and Advantages for Parallel Computation

Graphics Processing Units (GPUs) were initially designed for rendering images and video, but their architecture makes them exceptionally well-suited for parallel computation. Unlike CPUs, which have a few cores optimized for sequential processing, GPUs consist of thousands of smaller, efficient cores designed to handle multiple tasks simultaneously. This parallelism allows GPUs to perform many calculations concurrently, providing significant speedups for tasks that can be parallelized.

Advantages of GPUs:

Massive Parallelism: GPUs can execute thousands of threads in parallel, making them ideal for tasks with a high degree of parallelism, such as sequence alignment.

High Throughput: The ability to process many operations simultaneously results in high computational throughput, reducing the time required for data-intensive tasks.

Efficient Memory Management: GPUs are equipped with high-bandwidth memory and advanced memory management techniques, which can alleviate the memory bottlenecks encountered by CPUs.

Energy Efficiency: GPUs can achieve higher performance per watt compared to CPUs for parallelizable tasks, making them more energy-efficient for large-scale computations.

Programming Frameworks: Frameworks like CUDA (Compute Unified Device Architecture) and OpenCL (Open Computing Language) provide developers with tools to harness the power of GPUs for general-purpose computing, facilitating the development of GPU-accelerated bioinformatics algorithms.

By leveraging these advantages, GPU-accelerated genomic sequence alignment has the potential to overcome the limitations of traditional CPU-based approaches, enabling faster and more efficient analysis of large-scale genomic data.

III. GPU-Accelerated Sequence Alignment Algorithms

1. Overview of GPU-Accelerated Sequence Alignment Algorithms

Several GPU-accelerated sequence alignment algorithms have been developed to leverage the parallel processing capabilities of GPUs, providing substantial improvements in performance over traditional CPU-based methods. Notable examples include CUDA-BLAST and SOAP3-dp.

CUDA-BLAST: CUDA-BLAST is a GPU-accelerated version of the widely used BLAST algorithm. By implementing BLAST's core functions using CUDA (Compute Unified Device Architecture), CUDA-BLAST significantly reduces the time required for sequence alignment. CUDA-BLAST achieves this by parallelizing the search and alignment processes across multiple GPU cores, allowing for simultaneous processing of numerous sequence comparisons.

SOAP3-dp: SOAP3-dp is an advanced GPU-accelerated alignment tool designed for aligning short reads generated by high-throughput sequencing technologies. SOAP3-dp employs a hybrid approach that combines the Burrows-Wheeler Transform (BWT) and the FM-index to achieve fast and memory-efficient alignment. By utilizing the massive parallelism of GPUs, SOAP3-dp can handle large volumes of sequencing data with high speed and accuracy.

2. Description of Parallelization Strategies Used in GPU-Accelerated Algorithms

GPU-accelerated sequence alignment algorithms employ various parallelization strategies to maximize the computational power of GPUs. Key strategies include:

Data Parallelism: Data parallelism involves dividing the input data into smaller chunks and processing them simultaneously across multiple GPU threads. For instance, in CUDA-BLAST, the database sequences are divided into smaller segments, and each segment is aligned with the query sequence in parallel. This approach significantly speeds up the alignment process by taking advantage of the GPU's ability to handle numerous threads concurrently.

Task Parallelism: Task parallelism involves executing different tasks or stages of the alignment process in parallel. For example, SOAP3-dp splits the alignment process into several stages, such as searching, matching, and extending alignments. Each stage is then assigned to different GPU threads, allowing for concurrent execution and reducing overall alignment time.

Pipelining: Pipelining is a technique where different stages of the alignment process are executed in an overlapping manner. While one stage is being processed, the next stage begins execution on a different set of data. This continuous flow of data through the pipeline stages maximizes GPU utilization and minimizes idle time, leading to faster alignment.

Load Balancing: Efficient load balancing ensures that the computational workload is evenly distributed across GPU cores. Algorithms like SOAP3-dp dynamically adjust the workload assigned to each GPU thread based on its processing capabilities, preventing bottlenecks and ensuring optimal performance.

Memory Optimization: GPU-accelerated algorithms optimize memory usage to minimize data transfer between the CPU and GPU. Techniques such as shared memory and texture memory are utilized to store frequently accessed data, reducing the overhead of memory access and improving alignment speed.

3. Comparison of Performance Metrics Between GPU-Accelerated and CPU-Based Algorithms

Performance metrics such as alignment speed, throughput, and accuracy are crucial for evaluating the effectiveness of GPU-accelerated algorithms compared to their CPU-based counterparts. Several studies have demonstrated the superior performance of GPU-accelerated sequence alignment algorithms.

Alignment Speed: GPU-accelerated algorithms like CUDA-BLAST and SOAP3-dp can perform sequence alignments significantly faster than CPU-based methods. Benchmark tests have shown that CUDA-BLAST can achieve speedups of up to 10x compared to traditional BLAST, while SOAP3-dp can align short reads 20-50x faster than CPU-based Bowtie.

Throughput: The parallel processing capabilities of GPUs enable higher throughput, allowing for the simultaneous alignment of multiple sequences. This increased throughput is particularly beneficial for high-throughput sequencing data, where large volumes of reads need to be aligned quickly.

Accuracy: GPU-accelerated algorithms maintain high accuracy levels, comparable to or even exceeding those of CPU-based methods. The parallelization strategies employed ensure that the alignment results are consistent and reliable.

Resource Utilization: GPU-accelerated algorithms make efficient use of computational resources, resulting in lower power consumption and better performance per watt compared to CPU-based methods. This energy efficiency is advantageous for large-scale genomic studies, reducing operational costs and environmental impact.

IV. Case Studies and Applications

1. Case Studies Highlighting the Use of GPU-Accelerated Alignment in Bioinformatics Research

Case Study 1: Metagenomic Analysis

A study conducted by researchers at the University of California employed GPU-accelerated sequence alignment to analyze metagenomic samples from various environments, including soil,

ocean water, and the human gut. Traditional CPU-based methods were unable to process the vast amount of sequencing data within a reasonable timeframe. By utilizing CUDA-BLAST, the researchers achieved a tenfold reduction in alignment time, enabling the rapid identification of microbial species and their functional roles in different ecosystems. This acceleration facilitated timely and detailed insights into microbial diversity and community dynamics, contributing to a better understanding of environmental and human health.

Case Study 2: Cancer Genomics

In a project aimed at identifying genetic mutations associated with various cancers, a team at the National Cancer Institute implemented SOAP3-dp to align short reads obtained from cancer patient samples. The GPU-accelerated alignment allowed the team to process large datasets efficiently, significantly reducing the time required for alignment from several days to just a few hours. This rapid processing enabled the timely detection of mutations and their potential impacts on cancer progression and treatment response, highlighting the importance of GPU-accelerated alignment in personalized oncology.

Case Study 3: Comparative Genomics

Researchers at the European Bioinformatics Institute utilized GPU-accelerated sequence alignment to compare the genomes of multiple plant species, aiming to identify conserved and divergent genomic regions. By employing SOAP3-dp, the team achieved high-throughput alignment of large genomic sequences, which was previously impractical with CPU-based methods. The accelerated alignment process allowed for comprehensive comparative analysis, providing valuable insights into plant evolution, gene function, and adaptation mechanisms.

2. Examples of Applications

Metagenomics

Metagenomics involves the study of genetic material recovered directly from environmental samples. GPU-accelerated sequence alignment is particularly beneficial for metagenomic studies, where researchers need to process vast amounts of sequencing data to identify and characterize microbial communities. Accelerated alignment enables faster and more efficient analysis, facilitating the exploration of microbial diversity, ecosystem functions, and interactions between different microbial species. This has important implications for environmental monitoring, biotechnology, and human health.

Personalized Medicine

In personalized medicine, genetic information is used to tailor medical treatments to individual patients. GPU-accelerated sequence alignment plays a crucial role in analyzing patient genomes rapidly and accurately, identifying genetic variations associated with disease susceptibility and treatment response. By accelerating the alignment process, healthcare providers can deliver timely and precise diagnoses, optimize treatment plans, and improve patient outcomes.

Applications include identifying genetic markers for cancer, rare genetic disorders, and pharmacogenomics.

Evolutionary Studies

Evolutionary studies involve comparing the genomes of different species to understand their evolutionary relationships and the genetic basis of adaptation. GPU-accelerated sequence alignment allows researchers to handle large-scale genomic data efficiently, enabling comprehensive comparisons of multiple genomes. This facilitates the identification of conserved and divergent regions, evolutionary events such as gene duplications and horizontal gene transfer, and the elucidation of phylogenetic relationships. Accelerated alignment contributes to a deeper understanding of evolutionary processes and biodiversity.

Microbiome Research

Microbiome research focuses on the complex communities of microorganisms residing in various environments, including the human body. GPU-accelerated sequence alignment enables the rapid processing of sequencing data from microbiome samples, allowing researchers to identify microbial species and their functional roles. This is essential for understanding the impact of the microbiome on health and disease, developing probiotic therapies, and exploring the interactions between the microbiome and host.

Drug Discovery

In drug discovery, identifying genetic variations and understanding their functional implications is crucial for developing targeted therapies. GPU-accelerated sequence alignment aids in the rapid analysis of genomic data from patient cohorts and model organisms, facilitating the identification of drug targets and biomarkers. This accelerates the drug discovery pipeline, ena

V. Challenges and Future Directions

1. Challenges in Developing and Optimizing GPU-Accelerated Alignment Algorithms

Algorithm Complexity: Developing GPU-accelerated algorithms requires in-depth understanding of both the genomic alignment process and GPU architecture. This dual expertise is essential to effectively parallelize tasks and manage data flow between CPU and GPU.

Memory Management: Efficient memory utilization is crucial for optimizing performance. GPUs have limited memory compared to CPUs, and managing this memory efficiently to handle large genomic datasets poses a significant challenge. Techniques like memory coalescing and shared memory usage need careful implementation.

Load Balancing: Ensuring optimal load balancing across thousands of GPU cores is challenging. Uneven distribution of tasks can lead to some cores being idle while others are overloaded, resulting in suboptimal performance.

Data Transfer Overhead: Transferring data between CPU and GPU can introduce significant overhead. Minimizing these data transfers and optimizing the use of on-board GPU memory is critical for achieving high performance.

Algorithm Adaptation: Traditional sequence alignment algorithms must be adapted or redesigned to exploit GPU architecture effectively. This involves breaking down the algorithms into smaller, parallelizable tasks without compromising accuracy.

Development Tools and Debugging: Developing and debugging GPU-accelerated algorithms require specialized tools and frameworks like CUDA and OpenCL. These tools have a steep learning curve and can be complex to use effectively.

2. Future Directions for Enhancing GPU-Accelerated Genomic Sequence Alignment

Hybrid Architectures: Combining the strengths of both CPUs and GPUs in hybrid computing architectures can optimize performance. CPUs can handle sequential tasks and data management, while GPUs can perform parallelizable computations, leading to more efficient alignment algorithms.

Algorithm Optimization: Continuous optimization of existing algorithms is essential. This includes refining parallelization strategies, improving memory management techniques, and developing more sophisticated load balancing methods to fully exploit GPU capabilities.

Machine Learning Integration: Integrating machine learning techniques can enhance alignment accuracy and speed. For example, deep learning models can be trained to predict optimal alignment parameters or to identify regions of interest in sequences, reducing the computational load on GPUs.

Automated Toolchains: Developing automated toolchains that simplify the process of converting CPU-based algorithms to GPU-accelerated versions can lower the barrier to entry for researchers and developers, facilitating wider adoption of GPU acceleration.

Benchmarking and Standardization: Establishing standardized benchmarks for GPU-accelerated genomic sequence alignment can provide a consistent framework for evaluating and comparing different algorithms, driving improvements and innovation in the field.

Scalability and Cloud Computing: Leveraging cloud-based GPU resources can provide scalable solutions for large-scale genomic studies. Cloud platforms offer flexible access to high-performance computing resources, enabling researchers to process massive datasets efficiently.

3. Potential Impact of Emerging Technologies on Genomic Sequence Alignment

Quantum Computing: Quantum computing holds the promise of solving certain types of problems exponentially faster than classical computers. For genomic sequence alignment, quantum algorithms could potentially perform complex computations in a fraction of the time required by classical algorithms. While practical quantum computers are still in development,

ongoing research in quantum algorithms for sequence alignment could revolutionize the field in the future.

Neuromorphic Computing: Neuromorphic computing, which mimics the neural structure of the human brain, offers potential for highly efficient parallel processing. Neuromorphic hardware could be used to develop new types of alignment algorithms that are both fast and energy-efficient, further enhancing the capabilities of genomic analysis.

Advanced Memory Technologies: Emerging memory technologies like High Bandwidth Memory (HBM) and Non-Volatile Memory Express (NVMe) can provide faster data access and larger storage capacities. These advancements can help overcome current memory bottlenecks in GPU-accelerated alignment, improving performance and scalability.

Artificial Intelligence (AI) and Machine Learning (ML): AI and ML continue to evolve, offering new ways to optimize and enhance sequence alignment algorithms. Future advancements in AI-driven bioinformatics tools could provide more accurate and faster alignment solutions, integrating seamlessly with GPU-accelerated methods.

Edge Computing: With the rise of edge computing, processing data closer to the source (e.g., in sequencing machines or local servers) can reduce latency and improve the efficiency of genomic data analysis. Integrating GPU-accelerated algorithms into edge devices could enable real-time genomic analysis in clinical and research settings.

VI. Conclusion

Summary of the Benefits of GPU-Accelerated Genomic Sequence Alignment

GPU-accelerated genomic sequence alignment offers numerous advantages over traditional CPU-based methods. The primary benefits include:

Significant Speedup: By leveraging the parallel processing capabilities of GPUs, algorithms like CUDA-BLAST and SOAP3-dp achieve substantial reductions in alignment time. This speedup enables the rapid processing of large-scale genomic datasets, facilitating timely analysis and discovery.

High Throughput: GPU-accelerated alignment allows for the simultaneous processing of multiple sequences, significantly increasing throughput. This is particularly advantageous for high-throughput sequencing applications, where vast amounts of data need to be aligned quickly and efficiently.

Enhanced Scalability: GPU architectures are well-suited for handling the growing volume and complexity of genomic data. By distributing computational tasks across thousands of cores, GPU-accelerated algorithms can scale effectively, maintaining performance even as dataset sizes increase.

Energy Efficiency: GPUs offer higher performance per watt compared to CPUs for parallelizable tasks. This energy efficiency reduces operational costs and the environmental impact of large-scale genomic studies.

Maintained Accuracy: Despite the focus on speed and efficiency, GPU-accelerated algorithms maintain high levels of accuracy, ensuring reliable alignment results for various bioinformatics applications.

Importance of Continued Research and Development in This Field

Continued research and development in GPU-accelerated genomic sequence alignment are crucial for several reasons:

Algorithm Optimization: Ongoing efforts to optimize existing algorithms and develop new parallelization strategies will further enhance performance, making it possible to process even larger datasets more efficiently.

Integration with Emerging Technologies: As new technologies like quantum computing, neuromorphic computing, and advanced memory solutions emerge, integrating these advancements with GPU acceleration can unlock new levels of performance and capability in genomic analysis.

Tool Accessibility: Developing user-friendly tools and automated conversion processes will lower the barrier to entry for researchers and developers, facilitating broader adoption of GPU-accelerated methods in the bioinformatics community.

Addressing Challenges: Research focused on overcoming current challenges, such as efficient memory management and load balancing, will improve the robustness and reliability of GPU-accelerated algorithms, making them more versatile and widely applicable.

Final Thoughts on the Future of GPU-Accelerated Computing in Bioinformatics

The future of GPU-accelerated computing in bioinformatics looks promising. As sequencing technologies continue to advance and generate ever-larger datasets, the demand for efficient computational solutions will only grow. GPU acceleration has already demonstrated its potential to transform genomic sequence alignment, providing faster, more scalable, and energy-efficient solutions.

Looking ahead, the integration of GPU-accelerated methods with other emerging technologies will likely drive further innovation in bioinformatics. Quantum computing, AI-driven optimization, and advanced memory technologies are poised to complement GPU acceleration, pushing the boundaries of what is possible in genomic analysis.

Moreover, the development of hybrid computing architectures that combine the strengths of CPUs, GPUs, and other specialized hardware will enhance the versatility and performance of bioinformatics tools. These advancements will enable researchers to tackle increasingly complex

biological questions, leading to breakthroughs in personalized medicine, evolutionary studies, and environmental monitoring.

References

1. Elortza, F., Nühse, T. S., Foster, L. J., Stensballe, A., Peck, S. C., & Jensen, O. N. (2003). Proteomic Analysis of Glycosylphosphatidylinositol-anchored Membrane Proteins. *Molecular & Cellular Proteomics*, *2*(12), 1261–1270. https://doi.org/10.1074/mcp.m300079-mcp200

2. Sadasivan, H. (2023). *Accelerated Systems for Portable DNA Sequencing* (Doctoral dissertation).

3. Botello-Smith, W. M., Alsamarah, A., Chatterjee, P., Xie, C., Lacroix, J. J., Hao, J., & Luo, Y. (2017). Polymodal allosteric regulation of Type 1 Serine/Threonine Kinase Receptors via a conserved electrostatic lock. *PLOS Computational Biology/PLoS Computational Biology*, *13*(8), e1005711. https://doi.org/10.1371/journal.pcbi.1005711

4. Sadasivan, H., Channakeshava, P., & Srihari, P. (2020). Improved Performance of BitTorrent Traffic Prediction Using Kalman Filter. *arXiv preprint arXiv:2006.05540*.

5. Gharaibeh, A., & Ripeanu, M. (2010). *Size Matters: Space/Time Tradeoffs to Improve GPGPU Applications Performance*. https://doi.org/10.1109/sc.2010.51

6. Sankar S, H., Patni, A., Mulleti, S., & Seelamantula, C. S. (2020). Digitization of electrocardiogram using bilateral filtering. *bioRxiv*, 2020-05.

7. Harris, S. E. (2003). Transcriptional regulation of BMP-2 activated genes in osteoblasts using gene expression microarray analysis role of DLX2 and DLX5 transcription factors. *Frontiers in Bioscience*, *8*(6), s1249-1265. https://doi.org/10.2741/1170

8. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., & Hartl, F. U. (2013). Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annual Review of Biochemistry*, *82*(1), 323–355. https://doi.org/10.1146/annurev-biochem-060208-092442

9. Sankar, S. H., Jayadev, K., Suraj, B., & Aparna, P. (2016, November). A comprehensive solution to road traffic accident detection and ambulance management. In *2016 International Conference on Advances in Electrical, Electronic and Systems Engineering (ICAEES)* (pp. 43-47). IEEE.

10. Li, S., Park, Y., Duraisingham, S., Strobel, F. H., Khan, N., Soltow, Q. A., Jones, D. P., & Pulendran, B. (2013). Predicting Network Activity from High Throughput Metabolomics. *PLOS Computational Biology/PLoS Computational Biology*, *9*(7), e1003123. https://doi.org/10.1371/journal.pcbi.1003123

11. Liu, N. P., Hemani, A., & Paul, K. (2011). *A Reconfigurable Processor for Phylogenetic Inference*. https://doi.org/10.1109/vlsid.2011.74

12. Liu, P., Ebrahim, F. O., Hemani, A., & Paul, K. (2011). *A Coarse-Grained Reconfigurable Processor for Sequencing and Phylogenetic Algorithms in Bioinformatics*. https://doi.org/10.1109/reconfig.2011.1

13. Majumder, T., Pande, P. P., & Kalyanaraman, A. (2014). Hardware Accelerators in Computational Biology: Application, Potential, and Challenges. *IEEE Design & Test*, *31*(1), 8–18. https://doi.org/10.1109/mdat.2013.2290118

14. Majumder, T., Pande, P. P., & Kalyanaraman, A. (2015). On-Chip Network-Enabled Many-Core Architectures for Computational Biology Applications. *Design, Automation &Amp; Test in Europe Conference &Amp; Exhibition (DATE), 2015*. https://doi.org/10.7873/date.2015.1128

15. Özdemir, B. C., Pentcheva-Hoang, T., Carstens, J. L., Zheng, X., Wu, C. C., Simpson, T. R., Laklai, H., Sugimoto, H., Kahlert, C., Novitskiy, S. V., De Jesus-Acosta, A., Sharma, P., Heidari, P., Mahmood, U., Chin, L., Moses, H. L., Weaver, V. M., Maitra, A., Allison, J. P., . . . Kalluri, R. (2014). Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival. *Cancer Cell*, *25*(6), 719–734. https://doi.org/10.1016/j.ccr.2014.04.005

16. Qiu, Z., Cheng, Q., Song, J., Tang, Y., & Ma, C. (2016). Application of Machine Learning-Based Classification to Genomic Selection and Performance Improvement. In *Lecture notes in computer science* (pp. 412–421). https://doi.org/10.1007/978-3-319-42291-6_41

17. Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, *21*(2), 110–124. https://doi.org/10.1016/j.tplants.2015.10.015

18. Stamatakis, A., Ott, M., & Ludwig, T. (2005). RAxML-OMP: An Efficient Program for Phylogenetic Inference on SMPs. In *Lecture notes in computer science* (pp. 288–302). https://doi.org/10.1007/11535294_25

19. Wang, L., Gu, Q., Zheng, X., Ye, J., Liu, Z., Li, J., Hu, X., Hagler, A., & Xu, J. (2013). Discovery of New Selective Human Aldose Reductase Inhibitors through Virtual Screening Multiple Binding Pocket Conformations. *Journal of Chemical Information and Modeling*, *53*(9), 2409–2422. https://doi.org/10.1021/ci400322j

20. Zheng, J. X., Li, Y., Ding, Y. H., Liu, J. J., Zhang, M. J., Dong, M. Q., Wang, H. W., & Yu, L. (2017). Architecture of the ATG2B-WDR45 complex and an aromatic Y/HF motif crucial for complex formation. *Autophagy*, *13*(11), 1870–1883. https://doi.org/10.1080/15548627.2017.1359381

21. Yang, J., Gupta, V., Carroll, K. S., & Liebler, D. C. (2014). Site-specific mapping and quantification of protein S-sulphenylation in cells. *Nature Communications*, *5*(1). https://doi.org/10.1038/ncomms5776